Fine-scale oceanographic processes shape marine biodiversity patterns in the Galápagos Islands

Luke E Holman [1,2*]

Diana A. Pazmiño [3,4]

Shyam Gopalakrishnan [5]

Alexander Forryan [2]

Alex R. Hearn [3,4]

Alberto C. Naveira-Garabato [2]

Marc Rius [6,7]

[1] Section for Molecular Ecology and Evolution, Faculty of Health and Medical Sciences, Globe Institute, University of Copenhagen, Copenhagen, Denmark

[2] School of Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, United Kingdom

[3] Galápagos Science Center, Universidad San Francisco de Quito USFQ, Isla San Cristóbal, Galápagos 200150, Ecuador

[4] Migramar, Bodega Bay, California USA

[5] Section for Hologenomics, Faculty of Health and Medical Sciences, Globe Institute, University of Copenhagen, Copenhagen, Denmark

[6] Centre for Advanced Studies of Blanes (CEAB, CSIC), Accés a la Cala Sant Francesc 14, 17300 Blanes, Spain

[7] Department of Zoology, Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, Auckland Park, South Africa

*Corresponding author

lukeearlholman@gmail.com

Keywords

Marine biodiversity, ecological processes, metabarcoding, community structure, Galápagos Islands, oceanographic resistance

Abstract

Uncovering the drivers that shape biodiversity patterns is critical to understand fundamental ecological and evolutionary processes, but also to assist biodiversity managers and conservation agencies. Despite evidence that biodiversity composition is influenced by processes at different spatial scales, little is known about the role of fine-scale oceanographic processes in controlling marine biodiversity patterns. This is particularly important in biodiversity hotspot regions, where small changes in local conditions may facilitate introductions of novel species, local extirpation, or even extinction. Here, we conducted oceanographic modelling and environmental DNA (eDNA) metabarcoding to investigate how fine-scale oceanographic processes shape marine biogeographic patterns across the Galápagos Islands. We found that eDNA data confirmed previously reported biogeographic regionalization, and demonstrated significant differences in community structure across the highly diverse oceanographic seascape of the Galápagos Islands. We then tested the effect of local current systems with a novel metric, termed oceanographic resistance, measuring the cumulative seawater flow resistance between pairs of geographic sites. Oceanographic resistance explained a significant proportion of variation in eDNA-measured beta dissimilarity between sites (2.0% of total), comparable in influence to some of the most important abiotic drivers, such as temperature (2.9%) and geographic distance between sites (11.5%). This indicates that oceanographic resistance can be a useful metric to understand the effects of current systems on marine biota. Taken together, our results indicate that marine communities are particularly sensitive to changes in local current systems, and suggest that fine-scale oceanographic processes may have an underappreciated role in structuring marine communities globally.

Main

Spatial patterns of marine biodiversity are profoundly influenced by physical factors, such as oceanic currents and geographic barriers, with direct consequences on species distributions and community structure (1–3). Indeed, research has shown that planktonic communities increase in similarity proportionally to oceanographic connectivity (4–6), with distance travelled along currents having the same effect on beta diversity as geographic distance on land (so called distance-decay relationships) (7, 8). Despite decades of research into how oceanography shapes population connectivity, there are very few studies that explore the importance of local or finescale changes in ocean currents relative to other variables (such as temperature) in controlling marine biodiversity patterns (9, 10). Furthermore, we currently lack insight into whether submesoscale (horizontal scales < 100 km) ocean currents structure plankton biodiversity (11), or whether such currents affect free-swimming non-planktonic (nektonic) communities. This is particularly important in biodiversity hotspot regions, where even small changes in environmental conditions may lead to substantial conservation challenges.

The dispersal of holoplanktonic organisms and early life-history stages of nektonic organisms is often defined by ocean currents, with many species completing their life cycle adrift in the ocean. Finescale ocean currents may shape the distribution of nektonic species because: (i) a substantial proportion of nektonic organisms have planktonic early-life history stages (12); (ii) plankton and nekton are tightly connected through food webs (13, 14); and (iii) nektonic organisms tend to track thermal optima in current systems (15). Conversely, the distribution of some nektonic species may not be related to currents, as such species can migrate thousands of kilometres moving across current systems (16) and optimise behaviour to exploit resources that are rarely affected by ocean circulation (17, 18). In order to accurately test how factors such as finescale currents affect marine biodiversity patterns across different spatial scales, high-resolution biodiversity data are required.

In recent years, the use of high-throughput sequencing to analyse fragments of DNA found in the environment (often called environmental DNA or eDNA) has become common practice, and is now an established approach for marine biodiversity monitoring and a reliable way of producing whole-community data (19, 20). Our understanding of marine biodiversity is being revolutionised through eDNA surveys, with research uncovering previously undocumented global patterns (2, 5), revealing previously undescribed taxonomy (21) and, most recently, reconstructing long-dead marine taxa and biodiversity from ancient eDNA (22, 23). Despite all these advances, marine eDNA studies rarely integrate ocean circulation into their analyses (5, 24), and studies in community ecology have only explored the link between eDNA patterns and ocean currents with a relatively small subset of taxa (9, 10). There is therefore a pressing need to understand the potential role of ocean flows on biodiversity patterns considering a wide array of both planktonic and non-planktonic organisms.

Here, we elucidate the effect that ocean currents have on marine community structure across the waters surrounding the iconic Galápagos islands. We first use eDNA metabarcoding of seawater samples collected from across the archipelago to detect spatial patterns of fish and elasmobranch

biodiversity. Subsequently, we model the ocean circulation at high (submesoscale-permitting) resolution and infer the effect of eDNA decay to better understand the detected patterns of nektonic biodiversity. Finally, we develop a metric that describes local current systems from ocean model-generated data, motivated by the omission of ocean flow pathways in geographic distance-based metrics. We use this new metric to assess the relationships among ocean currents, a proxy for abiotic conditions (ocean temperature), and community dissimilarity.

Results

Galápagos fish biodiversity

Metabarcoding of eDNA water samples collected from sites across the Galápagos (Fig. 1a) produced a fish (teleost and elasmobranch) dataset containing 551 amplicon sequence variants (ASVs) of which 66 could be assigned to species level, 216 to Genus, 167 to Family, and 99 above Family level. Read numbers and diversity in negative control samples were typical for eDNA metabarcoding datasets (20) (full details provided in Supplementary Information 1).

Fish communities clustered in the nMDS ordination (Fig. 1c) according to previously reported bioregions (25). Specifically, the Western and Elizabeth bioregions appeared to cluster within each other, and were separated from the Northern and Central South-eastern bioregions. Roca redonda was a site not surveyed in Edgar et al. (25) (See Fig. S1), and clustered (top right of Fig. 1c) with sites from the Northern bioregion, and not with those from the Western bioregion as predicted by previous work (25).

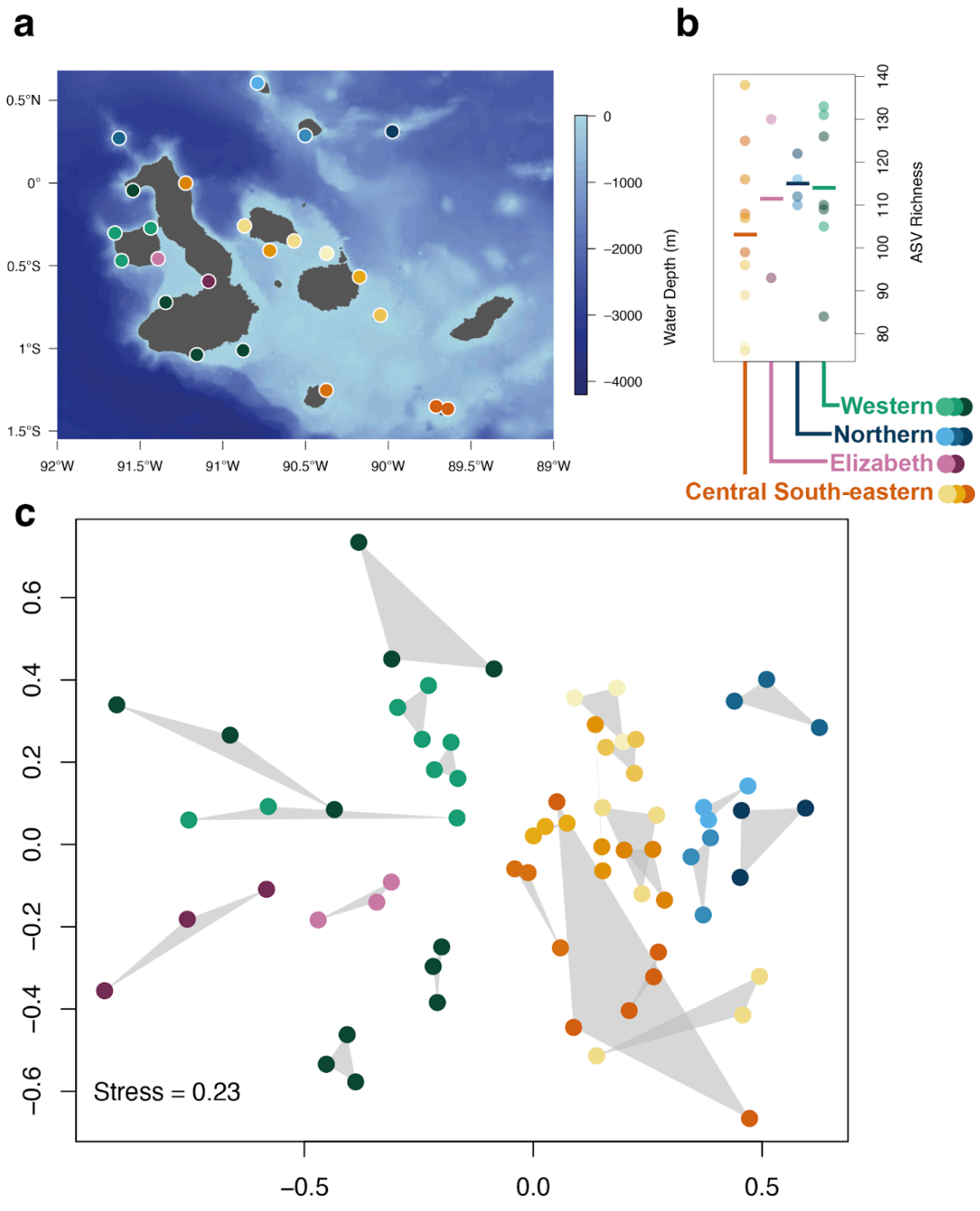

Fig. 1. *a) Map of the Galápagos islands, with sampling sites marked (dots) and depth indicated by blue colour gradient. b) ASV richness across the sampling sites grouped by the four main bioregions and averaged over field replicates, with the mean value indicated by a solid horizontal line. c) Non-metric multidimensional scaling based on Jaccard dissimilarity of community composition among sampling sites. Each point represents a single field replicate, with the three replicates per site joined by a grey convex hull. In all plots, point colour indicates bioregions from (25) as indicated in b).*

There was a statistically significant overall difference among bioregions (PERMANOVA $F_{3,19} = 2.08$, $p < 0.001$), with pairwise tests showing significant results ($p < 0.01$) among all bioregions except for the Elizabeth bioregion, which was not significantly ($p > 0.05$) different to any other bioregion (full model outputs in Table S1). Pairwise tests for significant differences in multivariate dispersion between bioregions (PERMDISP procedure) indicated that only the Elizabeth bioregion had significantly different multivariate dispersion compared to the Central South-eastern and Western bioregions ($p < 0.01$ in both cases, see Table S1 for full model output). A one-way ANOVA indicated no significant difference in mean ASV richness among bioregions (Fig. 1b) ($F_{3,19} = 0.72$, $p > 0.05$).

Finescale ocean currents influence local fish biodiversity

We found a positive relationship (distance-decay) between eDNA-measured site dissimilarity and geographic distance between pairs of sites (Fig. 2). To quantify the current faced by marine organisms travelling through the ocean, we calculated a novel metric that we termed oceanographic resistance. This metric is computed for pairs of sites, and is positive when the average flow along a given path in the ocean is in the same direction of travel, and negative when the average flow is against the direction of travel. We parameterized this metric using the average horizontal flow field values in a realistic, observationally ground-truthed, submesoscale-permitting ocean circulation model from the eDNA sampling month (26), also extracting the mean temperature for the month of eDNA sampling for each site in the model. There was a significant relationship between site dissimilarity and both geographic distance, temperature difference and oceanographic resistance ($F_{3,503} = 32.8$, $p < 0.01$ for all parameters). Geographic distance, temperature difference and oceanographic resistance explained 11.5%, 2.9% and 2.0% of the variation in the site dissimilarity index, respectively.

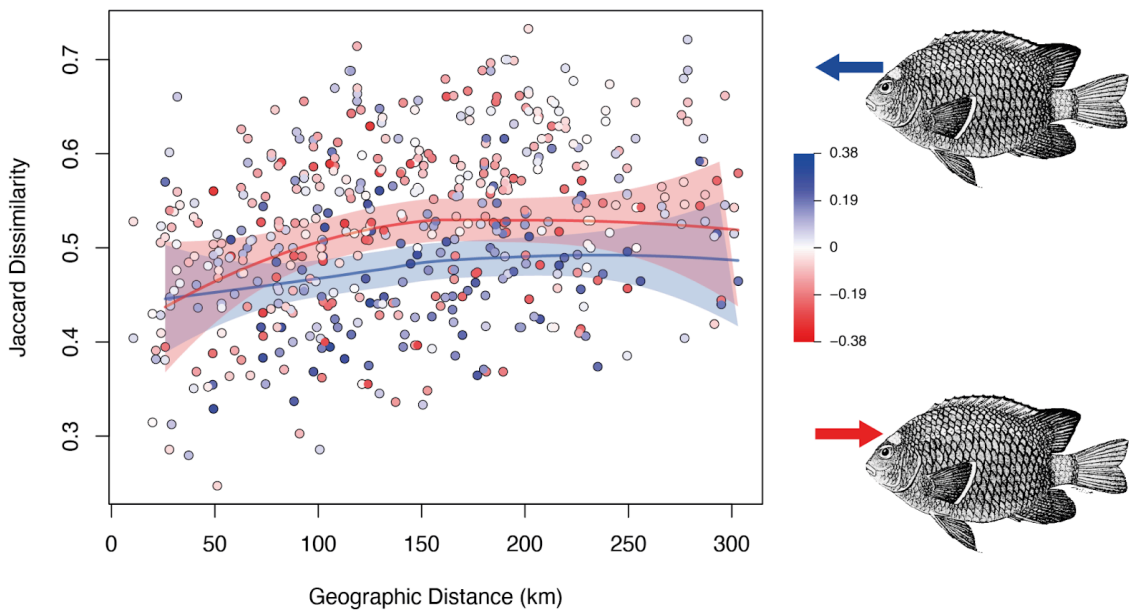

Fig. 2. *Modified asymmetric Jaccard dissimilarity for each pair of sites, displayed against geographic distance measured in km. Each point is coloured according to the oceanographic resistance between pairs of sites; point colour indicates oceanographic resistance with scale shown on the left, measured in m* s^{-1}*. Loess smoothed fit lines for data below the* 20th *percentile and above the* 80th *percentile of oceanographic resistance are shown as red and blue lines respectively, with shading indicating the 95% confidence interval of the fit. Fish illustrations(27) on the right denote the direction of average current flow for highly positive (blue) and highly negative (red) resistance.*

Across the distance-decay relationship when oceanographic resistance is negative, sites were on average more dissimilar than in cases where oceanographic resistance was positive (Fig. 2). A similar effect was observed with the temperature data, with greater dissimilarity between sites on average when temperature change between sites was positive (Fig. S2). Additionally, and in order to evaluate if fish ASV richness could also be linked to other measures of ocean circulation, particle release experiments were conducted with the same ocean circulation model used to define our oceanographic resistance metric. Particles were released into the model from the sampling sites, and run back in time for 72 hours to estimate possible eDNA contributions for each sampling event. No significant ($p > 0.05$) relationship was found between ASV richness and all four calculated metrics of oceanographic spread (e.g., mean distance from release point) (Figure S3).

Discussion

We found heterogenous fish community structure across the Galápagos islands, with eDNA metabarcoding-measured beta diversity patterns principally agreeing with previously described bioregions (25). Remarkably, we found not only that variation in fish communities could be explained with the submesoscale flow data generated by our ocean circulation model, but that the proportion of variance explained by currents was similar to temperature, a well-known determinant of marine biodiversity (15, 28). Overall, these results help us to not only better understand fish communities in this unique archipelago, but also provide a novel method to investigate the role of finescale currents on ecosystems across the globe.

Given that previous work has described fish bioregionalization across the archipelago (25), it is unsurprising that eDNA metabarcoding provides similar evidence for fish biogeography. However, more broadly, these patterns underline the unique nature of the Galápagos, with unusually clear differences in communities across short (<200 km) geographic distances. Many eDNA surveys have found biogeographic regionalization, particularly changes in community structure (beta diversity) in marine ecosystems (28–32). However, other studies have shown that marine fish communities can also have homogenous community structure, even across large (>1000 km) distances (33–35). Collectively these investigations suggest that homogenous biogeographic structure should be our null hypothesis for communities of highly mobile marine organisms at local regions. An important novel piece of biogeographic evidence in our study is the unexpected grouping of Roca Redonda in the Northern bioregion, this should prompt further research to investigate the, here un-sampled, Far-Northern Islands (Darwin & Wolf) which may have unanticipated biodiversity, potentially requiring a change in bioregion designation and thus management strategy. Given the limited sampling of the Elizabeth bioregion, further work is required to understand how, and if, fish communities in this region differ from the surrounding Western bioregion.

Our analysis combining novel oceanographic modeling and eDNA metabarcoding data could only explain a small proportion of the total variation among sites using distance and temperature data (Fig. 2 and Fig. S2). Studies evaluating the explanatory power of a set of environmental and/or spatial predictors typically only describe a small fraction of the total beta diversity in marine communities (28, 36, 37). These findings are also reflected in meta-analyses across ecosystems, with much of the measured variation in communities remaining unexplained (38–40). Metacommunity theory predicts that ecological drift (stochastic demographic changes in species composition) is likely to occur under both neutral and selective population dynamics (41), and thus some variation in community composition will always be unexplained by environmental and spatial predictors. Furthermore, there is frequently a compromise between surveying across space and through time to capture community dynamics, with even the most comprehensive ocean surveys showing only a snapshot of temporal dynamics (5). Therefore, typical ecological datasets are unlikely to provide a complete explanation of community structure from characteristically recorded parameters.

Our analyses indicated that geographic distance, temperature and oceanographic resistance - a metric of the finescale ocean currents' propensity (or opposition) to connect spatially separate sites - were important explanatory variables describing patterns in beta diversity in Galápagos fishes. Distance-decay relationships (changing biotic composition across space) are well studied in marine ecosystems; thus an effect of geographic distance was expected (8). Similarly, temperature has been shown to be a key variable structuring communities of both fish (42) and other marine organisms (5, 43). Work exploring the effect of current systems on marine biodiversity has either combined geographic and current-based distance (9), or been limited to benthic marine organisms (10). In line with our findings, these studies do find an important role for finescale ocean currents in structuring marine communities. We show that oceanographic resistance contains unique explanatory information, demonstrating that the direction and magnitude of currents connecting sites can influence the composition of fish communities. Given the significant effect of human-induced climate change on ocean currents and mixing (44), work is urgently needed to assess how currents on such fine scales will affect biodiversity patterns in other taxa and ecosystems.

Methods

Study area

The Galápagos Archipelago is made up of 13 major islands ranging in isolation from 3-25 NM from their nearest neighbor, lying in the Eastern Tropical Pacific Ocean, approximately 500 NM west from mainland Ecuador. Previous diver-based rocky reef surveys of fish and macroinvertebrates in shallow coastal waters around the islands (25) revealed a marked bioregional signal across the archipelago, with a warm, far northern region around the remote islands of Darwin and Wolf; a warm northern region encompassing the islands of Pinta, Genovesa and Marchena; a cool western region around Fernandina and western Isabela, and a mixed region around the central islands (see Fig. S1). The area including the channel separating Isabela and Fernandina, and adjacent Elizabeth Bay, was sufficiently different from the western bioregion to merit its own status.

Sample collection

We collected seawater samples from 23 sites across the southern and central Galápagos Islands (Fig.1a) during September 2018 (see Table S2 for details). At each sampling point 1 L of seawater was collected from 30 cm below the surface with a clean Kemmerer water sampler and filtered through a 0.22 µm polyethersulfone Sterivex filter (Merck Millipore, Massachusetts USA) using a sterile 50 ml luer lock syringe. Additionally, 2 L of seawater were collected at the maximum depth of each site (ranging from 11.4 to 100 feet) and filtered using the same method, resulting in a total of 3 L of water per site. As metazoan diversity detected by eDNA varies across depth (45), this approach aimed to characterise total fish diversity at the site. To minimise contamination among samples, after every filtration we bleached (5% solution) the materials and filtered distilled water as a negative control. We added 2 ml of ATL Buffer (Qiagen) to each Sterivex filter to preserve eDNA and stored them at room temperature until further processing.

DNA extraction and library preparation

We used the dedicated low-DNA laboratory at the National Oceanography Centre, Southampton (United Kingdom) to conduct the DNA extraction. This laboratory was separated from facilities where PCR was performed. No high copy DNA template, cultures or PCR products were permitted in this laboratory. All laboratory surfaces, reused equipment and reagent packaging were thoroughly cleaned with 5% bleach solution. DNA was extracted following the SX^{CAPSULE} method from (46), with sample identifiers blinded before extraction to avoid unintentional human bias. The final DNA elution was performed with 200ml DNase free water and an additional re-elution was performed with the eluate. Marine eDNA samples can contain PCR inhibitors, which have a negative effect on species

260 detection sensitivity (47). We therefore tested for inhibition using a Primer Design Internal
261 Positive Control qPCR Kit (Southampton, United Kingdom) and Primer Design Precision
262 Plus Mastermix with 20 µl reactions containing 4 µl of eDNA for each sample under the
263 manufacturer recommended conditions. Inhibited samples were expected to have an
264 increase in Ct (cycle threshold) of >1 compared to the unspiked reaction. As inhibition was
265 detected in a fraction of sites, all samples were treated using the Zymo OneStep PCR
266 Inhibition Removal Kit (Zymo Research, Irvine, United States of America or USA) following
267 the manufacturer recommended protocol.
268 We used metabarcoding primers that targeted a variably sized (163-185 bp) fragment of the
269 mitochondrial 12S region (48). These primers consist of two sets, one targeting teleost fish,
270 and a second set targeting elasmobranchs (sharks and rays). The entire metabarcoding
271 PCR and library build was performed independently for these two primer sets.
272 Metabarcoding libraries were constructed using a 2-step method where an initial PCR
273 incorporates an adaptor sequence onto the 5' end of the primers that serves as the target for
274 a second PCR that incorporates index sequences for demultiplexing and Illumina
275 sequencing adaptors (following Holman et al. 2021). For each set of primers PCR reactions
276 were conducted in 20 µl volumes consisting of 10 µl AmpliTaq Gold 360 mastermix (Agilent
277 Biosystems, Waltham, USA), 1.6 µl of primers (5 µM per primer) and 2 µl of undiluted
278 cleaned eDNA template. The reaction proceeded with an initial hold at 95°C for 10 minutes
279 followed by twenty cycles of 95°C for 30 seconds, 59°C for 30 seconds and 72°C for 60
280 seconds, a terminal hold at 72°C was conducted for 10 minutes. As the number of PCR
281 replicates per sample increases the diversity detected (49), we conducted eight independent
282 replicate reactions per water sample that were then pooled for bead cleaning and indexing.
283 These pools were cleaned using Beckman Coulter (Brea, USA) AMPure XP beads, for each
284 160 µl pool (eight 20 µl reactions), 128 µl of beads were added and the manufacturer
285 recommended protocol was followed with a final elution of DNA into 20 µl of 10mM Tris-HCl
286 buffer (pH 8.5). The second PCR was conducted in 20 µl volumes consisting of 10 µl
287 AmpliTaq Gold 360 mastermix (Agilent Biosystems, Waltham, USA), 1.0 µl of primers (10µM
288 per primer), and 5µl of bead-cleaned first PCR product. The reaction proceeded with an
289 initial hold at 95°C for 10 minutes followed by fifteen cycles of 95°C for 30 seconds, 55°C for
290 30 seconds and 72°C for 60 seconds, a terminal hold at 72°C was conducted for 10 minutes.
291 The product was then bead cleaned as above with 16 µl of beads in each 20µl reaction.
292 Libraries were then individually quantified using the New England Biolabs (Ipswich, United
293 States) NEBNext Library Quant Kit and pooled at equal molarity into two libraries, one for
294 each initial primer set. These two libraries were diluted to 4 nM, spiked with 5% PhiX for
295 diversity and sequenced in two independent runs of a Illumina (San Diego, United States)
296 MiSeq instrument using a V3 2x300 bp kit. Negative controls from sampling, DNA extraction

297 and PCR one and two blanks (RT-PCR grade water) were all amplified, pooled and
298 sequenced along with experimental samples.
299
300 *Bioinformatic analyses*
301 Raw sequences were demultiplexed using the GenerateFastQ (v.2.0.0.9) module on the
302 MiSeq control software (v.3.0.0.105) under default settings. Primers were then trimmed from
303 both paired reads, ensuring both the forward and reverse primer was present in each read
304 pair using Cutadapt (50) (v.3.2). As the sequencing length covered the entire target
305 amplicon, the reverse complement of each primer was also trimmed using Cutadapt from the
306 3'-end of each read pair. Following primer trimming reads were denoised into ASVs using
307 the DADA2 pipeline (v1.16.0) (51) in R (52) (v.4.0.3) under default parameters unless
308 detailed below. The *filterAndTrim* function was conducted using a maxEE value of 1 and
309 truncLen value of 120 bp for both read pairs. After the generation of an ASV table the data
310 was curated by running lulu (v.0.1.0) (53) under the default parameters. Each independently
311 sequenced dataset was then cleaned separately as follows using R. Positive observations
312 were discarded if they had fewer raw reads than the sum of all reads found in the negative
313 control samples for each ASV, or if they had fewer than three reads. ASV by sample tables
314 were then transformed into proportional abundance per sample and data from identical
315 ASVs was merged using the collapseNoMismatch function in DADA2. It is commonplace to
316 multiplex the two MiFish primer sets (elasmobranch and teleost) during PCR and treat them
317 as a single marker (48, 49), because they differ by only three nucleotides across the forward
318 and reverse primers, they amplify many species in common. However, we instead chose to
319 increase the sequencing output per sample and conservatively merged these primer sets
320 bioinformatically as above. ASVs were then searched against the entire NCBI *nt* database
321 (updated on 01-02-2021) using a BLAST+ (v.2.11.0) nucleotide search with
322 *-num_alignments* set to 200. These alignments were then filtered using the R script
323 ParseTaxonomy (DOI:10.5281/zenodo.4564075) that was previously developed (28) to
324 generate lowest-common-ancestor assignments in the case of multiple matches. Initial
325 analyses revealed some errors in the assignments, likely due to missing taxa in databases,
326 so all ASV assignments were curated using the online NCBI blastn portal (accessed
327 March-August 2022). Erroneous sequences were identified as having no match to any
328 nucleotide or protein (using a blastx search) 12S sequence, all such sequences were
329 discarded. ASVs matching domestic animals (cow, dog, chicken etc.) or human DNA were
330 removed from the main dataset. Finally, ASVs with an unambiguous species assignment
331 (>99% sequence similarity across the whole sequence, matches to other species in the
332 genus >1% sequence similarity from the proposed assignment) were merged.
333

Oceanographic analyses

Particle tracking simulations were conducted using a realistic, observational ground-truthed, previously described oceanographic model (26) constructed using MITgcm (54) with bathymetry from General Bathymetric Chart of the Oceans (GEBCO_14) Grid. Model grid resolution was initially 4 km in the horizontal between ± 5º latitude stretching out to ~12 km in latitude at the model boundaries, with 840 grid points in X and 600 in Y and a grid origin at 17.8ºS, 105ºW. The vertical grid comprised 75 depth levels, with vertical resolution varying with depth from 5 m over the first 50 m, 9.8 m to 164 m depth, and 13.7 m to 315 m depth, and a maximum cell height of 556 m below 3000 m. This model was run with three completely open boundaries (North, South and West), using periodic boundary forcing for temperature, salinity and velocity fields and a 15-grid box thick sponge layer for velocity. Initial conditions and monthly boundary forcing were taken from the Mercator Ocean reference model (https://www.mercator-ocean.fr/), a global ocean model based on 1/12 (0.083) degree NEMO (https://www.nemo-ocean.eu/).

Following the initial four km resolution model run, a smaller area encompassing the Galápagos Marine Reserve was modelled at 1 km horizontal resolution using the same vertical resolution as the 4 km model. The 1 km model has 630 grid points in X and 768 grid points in Y, with a grid origin at 3.1ºS 94.1ºW. Boundary forcing and initial conditions for the 1 km model were taken from the 4 km model.

Atmospheric forcing, wind stress and evaporation and precipitation for both models were taken from the ERA-Interim (55) reanalysis at a 3-hour temporal resolution for all fields, and radiation (shortwave and longwave) forcing from Modern-Era Retrospective analysis for Research and Applications (v.2) (MERRA2 (56)) at hourly temporal resolution.

Particle tracing experiments were performed in the 1 km model using TRACMASS (57) to establish the likely origin of waters at the sample sites. Particles were released five times for each sample site (2 days before to 2 days after) sampling, covering a horizontal area of ~ 4 km^2 around the site, from the surface to 20 m depth. These particles were then tracked backwards-in-time through the model flow field for 3 days. The final positions of particles from all releases were aggregated and normalised (as a fraction, where one is the sum of all particles released) and a spatial distribution for likely sample site water origin estimated. Four parameters describing the spread of the particles 48 hours before sampling were calculated. The direct line distance between the average latitude and longitude of the points from the release point, the mean distance of the particles from the mean latitude and longitude of the points, the surface area occupied by grid squares with greater than 0.01% of released particles, and the average of the individual particle direct line distances from the release point.

Calculating oceanographic resistance

The General Bathymetric Grid of the Oceans 2022 grid (58) was subset around the Galápagos Islands using the sf package (v.1.0.9.) (59) in R (v.4.2.2). This dataset contains seawater depth and coastline information at a resolution of 15 arcseconds. For each possible journey from every site to every other site the shortest path avoiding land masses was calculated using the *shortestPath* function in the *gdistance* R package (v.1.6) (60). These data are henceforth referred to as geographic distance.

To estimate the overall water resistance experienced by an agent travelling along the shortest path in the study area between two sites, taking into account ocean currents, we devised a metric that we refer to as oceanographic resistance, calculated as follows. For each site-to-site geographic distance, a point was extracted from along the path every 1 km using the *spsample* function from the *sf* R package. Northings were extracted from the 1 km model as mean monthly northwards velocity (m s^{-1} positive going north) and Eastings as mean monthly eastwards velocity (m s^{-1} positive going east) from the 1 km model for 2018. From these Northing and Easting values the resultant vector was calculated and represented by magnitude and azimuth degrees. The azimuth of the oceanographic current for each extracted point was then compared to the azimuth between the extracted point and the subsequent point along the path. The resultant angle indicates the difference between the direction of travel and the direction of the current, with for example, zero degrees indicating that the current and direction of travel are identical and 180 degrees indicating that the current and direction of travel are opposite. This comparison angle was then transformed using a cosine function to give a value of 1 and -1, respectively for the previous examples. The oceanographic resistance at the extracted point was calculated by multiplying the result of the cosine function by the magnitude of the current at the point. Finally, the oceanographic resistance for a given path was calculated as the mean of the oceanographic resistance of all selected points on the path between the start and end points. Oceanographic resistance is a mean value of a series of transformed vectors measured in m s^{-1}, and as such is a scalar measured in m s^{-1}.

Ecological analyses

All analyses were conducted in R (v.4.2.2) unless otherwise stated. Differences in mean ASV richness between bioregions were evaluated using a one-way ANOVA. Beta dissimilarity between sites was visualised with non-metric multidimensional scaling (nMDS) using a Jaccard dissimilarity, an index appropriate for testing biogeographical patterns (61), implemented with the metaMDS function from vegan (v.2.6-4) (62). All beta diversity analyses were conducted on averaged values among the three replicates per site. Subsequent statistical tests on bioregions followed the designation of (25) with one site

(Roca Redonda) in a previously unsurveyed region placed in a bioregion according to the results of the nMDS. Differences in within-bioregion multivariate dispersion were evaluated using the PERMDISP (63) procedure implemented in *betadisper* function from vegan, with post hoc testing of pairwise differences tested using the *TukeyHSD* function. Statistically significant differences between bioregions were evaluated using a PERMANOVA (64) on Jaccard dissimilarities implemented using the *adonis2* function in vegan. Pairwise PERMANOVA comparisons between bioregions were implemented using the *adonis.pair* function in the EcolUtils package (v.0.1) (65).

To test for possible correlations between ASV richness and particle spread, least-square regression models were implemented using the function *lm* with each of the particle spread statistics described above. Relationships between beta dissimilarity and particle spread characteristics were evaluated using a distance-based redundancy analysis implemented with the *dbrda* function from vegan and Jaccard dissimilarities.

In contrast to oceanographic resistance, Jaccard dissimilarity is symmetric considering the order of sites. For example, for a pair of sites, the oceanographic resistance defined above may differ depending on the direction of travel from site A to site B, while the Jaccard dissimilarity measures the difference between sites symmetrically. In order to test for an effect of oceanographic resistance on beta diversity a modified asymmetric Jaccard dissimilarity was implemented such that the order of the two sites supplied to the function (i.e. Site A to Site B / Site B to Site A) changed the output as below.

$$Jaccard(A, B) = \frac{A \cap B}{A \cup B}$$

$$Asymmetric\ Dissimilarity(A, B) = \frac{A \cap B}{A}$$

This modified dissimilarity measure can be interpreted as the dissimilarity between site A and site B considering only species present in site A. In other words, species not found in site A that are present in site B do not contribute to the dissimilarity index. Asymmetric dissimilarity was used as the dependent variable in a least-squares regression against geographic distance with an additive effect of oceanographic resistance with the *lm* function in R, values comparing sites to themselves were omitted before analysis. In order to evaluate the comparative effect of oceanographic resistance to other marine conditions we extracted average October 2018 mean sea temperature from the top 20 M of the model output for each site. These values were transformed into temperature differences between sites and incorporated in the above linear model as an additive effect. The *etasq* function from the heplots (v.1.4-2) was used to calculate the partial R^2 for each of the variables (66).

Acknowledgements

All authors thank the Galápagos National Park Directorate (GNPD), Universidad San Francisco de Quito (USFQ), and Galapagos Science Center (GSC) for logistic support for fieldwork, and the Ministry of Environment, Ecuador for granting a collection permit (MAE-DNB-CM-2016-0041). The work received support from a UK Royal Society Grant (CH160019) led by ANG. LH was supported by the Natural Environmental Research Council UK (grant number NE/L002531/1) and the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 856488). MR acknowledges support from the MARGECH project (PID2020-118550RB) of the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033). The authors thank Julián Regalado Pérez for discussions on oceanographic resistance.

Author Contributions

ANG initiated the project and secured funding, supported by DP, AH and MR. LH, DP and MR designed the field sampling and samples processing strategy. DP conducted the fieldwork. LH conducted the laboratory work, bioinformatics, ecological analyses and wrote the initial manuscript. AF and ACNG conducted oceanographic modelling analyses. All authors substantially contributed to further manuscript drafts and provided final approval for publication.

Competing Interests

The authors declare no competing interests.

Data & Code availability

The raw Illumina sequencing data are available from the European Nucleotide Archive under study accession number PRJEB55415. All other metadata, intermediate data and scripts are permanently archived at DOI: 10.5281/zenodo.10593433.

References

1. B. A. Ward, B. B. Cael, S. Collins, C. R. Young, Selective constraints on global plankton dispersal. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2007388118 (2021).

2. G. Sommeria-Klein, *et al.*, Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science* **374**, 594–599 (2021).

3. M. J. Costello, *et al.*, Marine biogeographic realms and species endemicity. *Nat. Commun.* **8**, 1057 (2017).

4. E. Villarino, *et al.*, Large-scale ocean connectivity and planktonic body size. *Nat. Commun.* **9**, 142 (2018).

5. D. J. Richter, *et al.*, Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *Elife* **11**, e78129 (2022).

6. J. Bendtsen, L. L. Sørensen, N. Daugbjerg, N. Lundholm, K. Richardson, Phytoplankton diversity explained by connectivity across a mesoscale frontal system in the open ocean. *Sci. Rep.* **13**, 12117 (2023).

7. D. R. Clark, G. J. C. Underwood, T. J. McGenity, A. J. Dumbrell, What drives study-dependent differences in distance–decay relationships of microbial communities? *Glob. Ecol. Biogeogr.* **30**, 811–825 (2021).

8. C. Graco-Roza, *et al.*, Distance decay 2.0 - A global synthesis of taxonomic and functional turnover in ecological communities. *Glob. Ecol. Biogeogr.* **31**, 1399–1421 (2022).

9. J. R. Watson, *et al.*, Currents connecting communities: nearshore community similarity and ocean circulation. *Ecology* **92**, 1193–1200 (2011).

10. A. Rattray, *et al.*, Geographic distance, water circulation and environmental conditions shape the biodiversity of Mediterranean rocky coasts. *Mar. Ecol. Prog. Ser.* **553**, 1–11 (2016).

11. M. Lévy, P. J. S. Franks, K. S. Smith, The role of submesoscale currents in structuring marine ecosystems. *Nat. Commun.* **9**, 4758 (2018).

12. K. O. Winemiller, K. A. Rose, Patterns of Life-History Diversification in North American Fishes: implications for Population Regulation. *Can. J. Fish. Aquat. Sci.* **49**, 2196–2218 (1992).

13. D. G. Boyce, K. T. Frank, W. C. Leggett, From mice to elephants: overturning the “one size fits all” paradigm in marine plankton food chains. *Ecol. Lett.* **18**, 504–515 (2015).

14. T. Fenchel, Marine Plankton Food Chains. *Annu. Rev. Ecol. Syst.* **19**, 19–38 (1988).

15. J. M. Sunday, A. E. Bates, N. K. Dulvy, Thermal tolerance and the global redistribution of animals. *Nat. Clim. Chang.* **2**, 686–690 (2012).

16. B. A. Block, *et al.*, Tracking apex marine predator movements in a dynamic ocean. *Nature* **475**, 86–90 (2011).

17. D. W. Sims, *et al.*, Scaling laws of marine predator search behaviour. *Nature* **451**, 1098–1102 (2008).

18. D. W. Sims, V. A. Quayle, Selective foraging behaviour of basking sharks on zooplankton in a small-scale front. *Nature* **393**, 460–464 (1998).

19. K. Deiner, *et al.*, Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol. Ecol.* **26**, 5872–5895 (2017).

20. M. Miya, Environmental DNA Metabarcoding: A Novel Method for Biodiversity Monitoring of Marine Fish Communities. *Ann. Rev. Mar. Sci.* **14**, 161–185 (2022).

21. M. Gaïa, *et al.*, Mirusviruses link herpesviruses to giant viruses. *Nature* **616**, 783–789 (2023).

22. K. H. Kjær, *et al.*, A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature* **612**, 283–291 (2022).

23. H. H. Zimmermann, *et al.*, Marine ecosystem shifts with deglacial sea-ice loss inferred from ancient DNA shotgun sequencing. *Nat. Commun.* **14**, 1650 (2023).

24. E. A. Andruszkiewicz, *et al.*, Modeling Environmental DNA Transport in the Coastal Ocean Using Lagrangian Particle Tracking. *Frontiers in Marine Science* **6**, 477 (2019).

25. G. J. Edgar, S. Banks, J. M. Farina, M. Calvopina, C. Martinez, Regional biogeography of shallow reef fish and macro-invertebrate communities in the Galapagos archipelago. *J. Biogeogr.* **31**, 1107–1124 (2004).

26. A. Forryan, A. C. Naveira Garabato, C. Vic, A. J. G. Nurser, A. R. Hearn, Galápagos upwelling driven by localized wind-front interactions. *Sci. Rep.* **11**, 1277 (2021).

27. W. Fischer, *Guia FAO para la identificación de especies para los fines de la pesca: Vertebrados, parte 2* (FAO, 1995).

28. L. E. Holman, *et al.*, Animals, protists and bacteria share marine biogeographic patterns. *Nat Ecol Evol* **5**, 738–746 (2021).

29. J. D. DiBattista, *et al.*, Environmental DNA reveals a multi-taxa biogeographic break across the Arabian Sea and Sea of Oman. *Environ. DNA* **4**, 206–221 (2022).

30. K. West, *et al.*, Large-scale eDNA metabarcoding survey reveals marine biogeographic break and transitions over tropical north-western Australia. *Divers. Distrib.* **27**, 1942–1957 (2021).

31. O. N. Marwayana, Z. Gold, C. P. Meyer, P. H. Barber, Environmental DNA in a global biodiversity hotspot: Lessons from coral reef fish diversity across the Indonesian archipelago. *Environ. DNA* **4**, 222–238 (2022).

32. M. Kume, *et al.*, Factors structuring estuarine and coastal fish communities across Japan using environmental DNA metabarcoding. *Ecol. Indic.* **121**, 107216 (2021).

33. N. Fraija-Fernández, *et al.*, Marine water environmental DNA metabarcoding provides a comprehensive fish diversity assessment and reveals spatial patterns in a large oceanic area. *Ecol. Evol.* **10**, 7560–7584 (2020).

34. M. R. Jensen, *et al.*, Distinct latitudinal community patterns of Arctic marine vertebrates along the East Greenlandic coast detected by environmental DNA. *Divers. Distrib.* **29**, 316–334 (2023).

35. T. E. Berry, *et al.*, A 3-year plankton DNA metabarcoding survey reveals marine biodiversity patterns in Australian coastal waters. *Divers. Distrib.* **29**, 862–878 (2023).

36. S. C. Burgess, K. Osborne, M. J. Caley, Similar regional effects among local habitats on the structure of tropical reef fish and coral communities. *Glob. Ecol. Biogeogr.* **19**, 363–375 (2010).

37. J. Ramos Miranda, D. Mouillot, A. Sosa Lopez, T. Do Chi, D. Flores Hernández, How much variation can be explained by seasonal, spatial and environmental effects in nekton assemblages of the Terminos Lagoon? *Aquat. Conserv.* **18**, 508–517 (2008).

38. J. Soininen, Spatial structure in ecological communities – a quantitative analysis. *Oikos* **125**, 160–166 (2016).

39. J. Soininen, A quantitative analysis of species sorting across organisms and ecosystems. *Ecology* **95**, 3284–3292 (2014).

40. K. Cottenie, Integrating environmental and spatial processes in ecological community dynamics. *Ecol. Lett.* **8**, 1175–1182 (2005).

41. M. A. Leibold, J. M. Chase, *Metacommunity Ecology, Volume 59* (Princeton University Press, 2018).

42. M. Montanyès, B. Weigel, M. Lindegren, Community assembly processes and drivers shaping marine fish community structure in the North Sea. *Ecography* **2023** (2023).

43. D. P. Tittensor, *et al.*, Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101 (2010).

44. S. C. Doney, *et al.*, Climate change impacts on marine ecosystems. *Ann. Rev. Mar. Sci.* **4**, 11–37 (2012).

45. O. Canals, I. Mendibil, M. Santos, X. Irigoien, N. Rodríguez-Ezpeleta, Vertical stratification of environmental DNA in the open ocean captures ecological patterns and behavior of deep-sea fishes. *Limnol. Oceanogr. Lett.* **6**, 339–347 (2021).

46. J. Spens, A. R. Evans, D. Halfmaerten, Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods Ecol. Evol.* **8**, 635–645.

47. A. M. McKee, S. F. Spear, T. W. Pierson, The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biol. Conserv.* **183**, 70–76 (2015).

48. M. Miya, *et al.*, MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R Soc Open Sci* **2**, 150088 (2015).

49. H. Doi, *et al.*, Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. *Sci. Rep.* **9**, 3581 (2019).

50. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

51. B. J. Callahan, *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

52. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project.org/* (2023) (April 14, 2023).

53. T. G. Frøslev, *et al.*, Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.* **8**, 1188 (2017).

54. J. Marshall, A. Adcroft, C. Hill, L. Perelman, C. Heisey, A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.* **102**, 5753–5766 (1997).

55. D. P. Dee, *et al.*, The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.* **137**, 553–597 (2011).

56. Global Modeling and Assimilation Office, S. Pawson, MERRA-2 tavg1_2d_int_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Vertically Integrated Diagnostics V5.12.4 (2015) https:/doi.org/10.5067/Q5GVUVUIVGO7.

57. K. Döös, B. Jönsson, J. Kjellsson, Evaluation of oceanic and atmospheric trajectory schemes in the TRACMASS trajectory model v6.0. *Geosci. Model Dev.* **10**, 1733–1749 (2017).

58. GEBCO Bathymetric Compilation Group, The GEBCO_2022 Grid - a continuous terrain model of the global oceans and land (2022) https:/doi.org/10.5285/E0F0BB80-AB44-2739-E053-6C86ABC0289C.

59. E. Pebesma, Simple features for R: Standardized support for spatial vector data. *R J.* **10**, 439

(2018).

60. J. van Etten, R Package gdistance: Distances and Routes on Geographical Grids. *J. Stat. Softw.* **76**, 1–21 (2017).

61. H. Kreft, W. Jetz, A framework for delineating biogeographical regions based on species distributions. *J. Biogeogr.* **37**, 2029–2053 (2010).

62. J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, vegan: Community Ecology Package. R package version 2.6-4. 2023 (2023).

63. M. J. Anderson, Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**, 245–253 (2006).

64. M. J. Anderson, Permutational multivariate analysis of variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online*, 1–15 (2017).

65. G. Salazar, EcolUtils: Utilities for community ecology analysis. *R package* (2023).

66. J. Fox, M. Friendly, G. Monette, Heplots: Visualizing tests in multivariate linear models. *R package version 1.4-2* (2022).

Figures

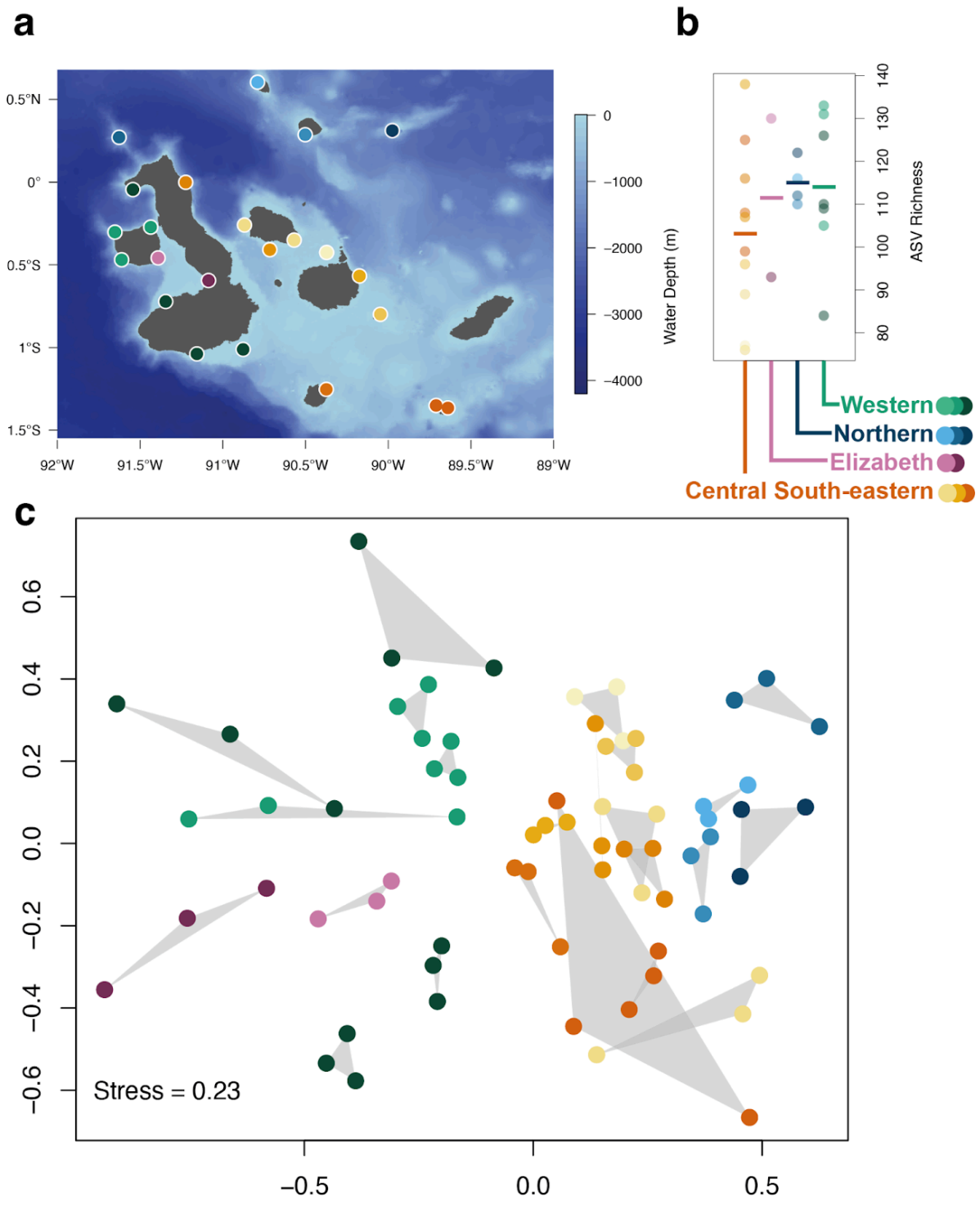

Figure 1. a) Map of the Galápagos islands, with sampling sites marked (dots) and depth indicated by blue colour gradient. b) ASV richness across the sampling sites grouped by the four main bioregions and averaged over field replicates, with the mean value indicated by a solid horizontal line. c) Non-metric multidimensional scaling based on Jaccard dissimilarity of community composition among sampling sites. Each point represents a single field replicate, with the three replicates per site joined by a grey convex hull. In all plots, point colour indicates bioregions from (25) as indicated in b).

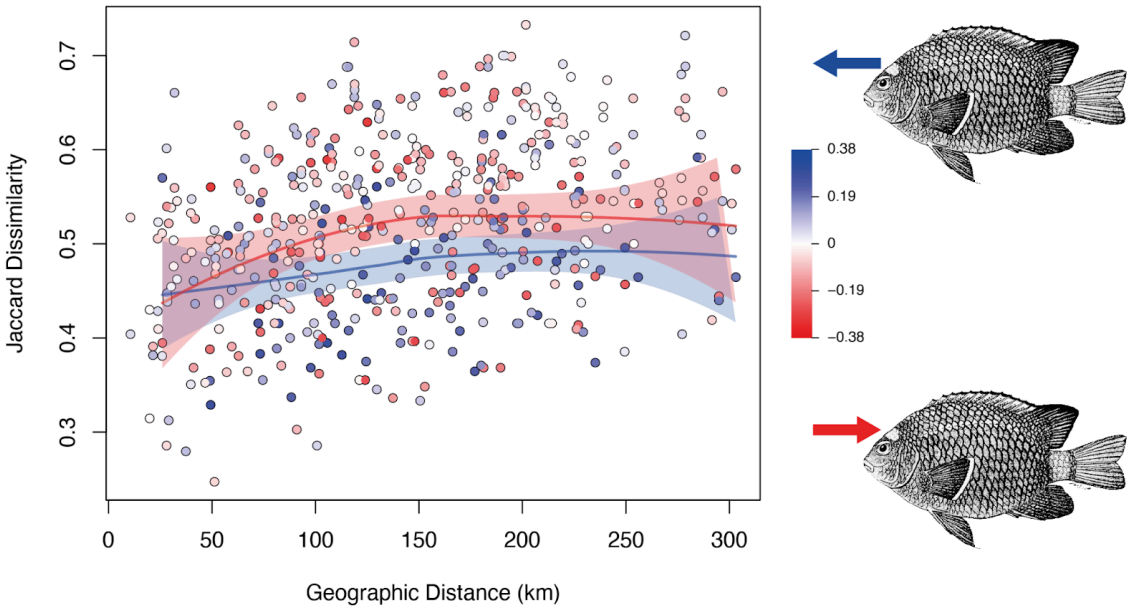

Figure 2. Modified asymmetric Jaccard dissimilarity for each pair of sites, displayed against geographic distance measured in km. Each point is coloured according to the oceanographic resistance between pairs of sites; point colour indicates oceanographic resistance with scale shown on the left, measured in ms^{-1}. Loess smoothed fit lines for data below the 20th percentile and above the 80th percentile of oceanographic resistance are shown as red and blue lines respectively, with shading indicating the 95% confidence interval of the fit. Fish illustrations on the right denote the direction of average current flow for highly positive (blue) and highly negative (red) resistance.